

The impact of mandatory policies on ETD acquisition

Arthur Sale

University of Tasmania

Arthur.Sale@utas.edu.au

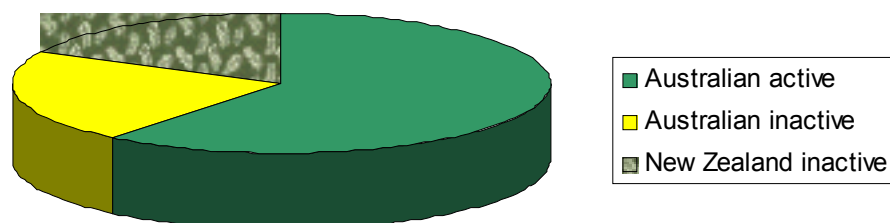
This paper analyzes the data now available in Australia's coordinated Electronic Theses and Dissertations gateway to show the impact of high-level institutional policy decisions on population of the individual repositories. The paper shows that just like research article repositories, voluntary ETD deposition results in repositories collecting only about 15% of the available theses, whereas mandatory policies are well accepted and cause deposit rates to rise towards 100%. Modeling of the PhD and Master process in Australia is also carried out to indicate the delays and liabilities to be expected if mandatory policies are applied only to newly enrolled candidates.

Participation and software

The Australian university system has a common gateway to *electronic theses and dissertations* (ETDs) produced in its universities. This gateway, the *Australian Digital Theses Program* (ADT Program, 2005), harvests from 28 (74%) of the 38 Australian universities (there are 29 repositories, but two of them belong to one university), and is in its turn harvested by Google, Yahoo! and scholarly search engines. The remaining 10 universities have not yet established local ETD repositories. Nevertheless, the Program has a potential coverage of 84% of Australian university research degrees, since many of the non-contributors are smaller universities without strong research programs.

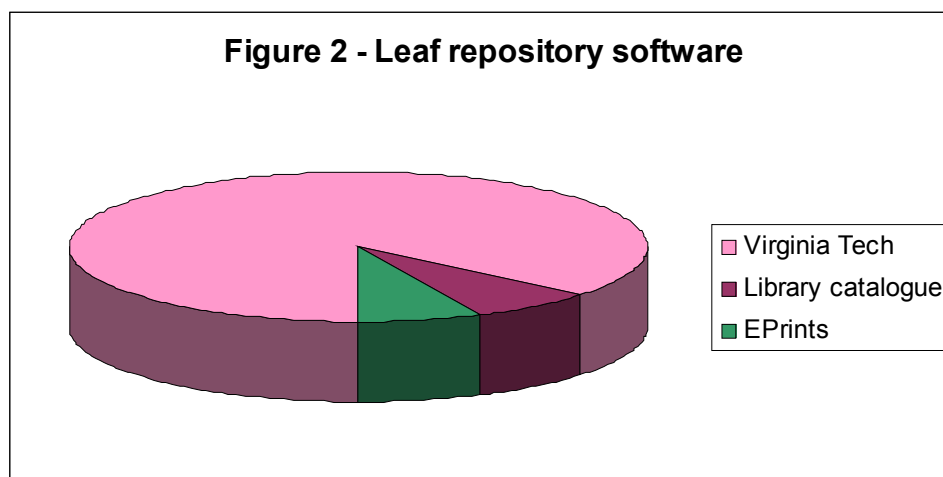
The participation rate must be regarded as a significant success by the ADT Program in a short time. An agreement has recently been reached with CONZUL (the *Committee of New Zealand University Libraries*) to have the eight New Zealand universities also contribute ETDs, but none have as yet established repositories. The current situation is shown in Figure 1. The ADT Program has been running since 2000; for a sufficiently long time that an analysis of its success can be undertaken, as a work in progress.

Figure 1 - ADT program participation



The gateway harvests metadata from (local) leaf repositories in the universities for its own database. Searchers who find the gateway metadata pages either through a search engine or via a direct search of the gateway are directed to the leaf repository, which is the only place that the 'full-text' is stored. While the ADT Program is oriented around text documents which are required to be in pdf format, a few theses have associated materials such as software, again only stored in the leaf repositories.

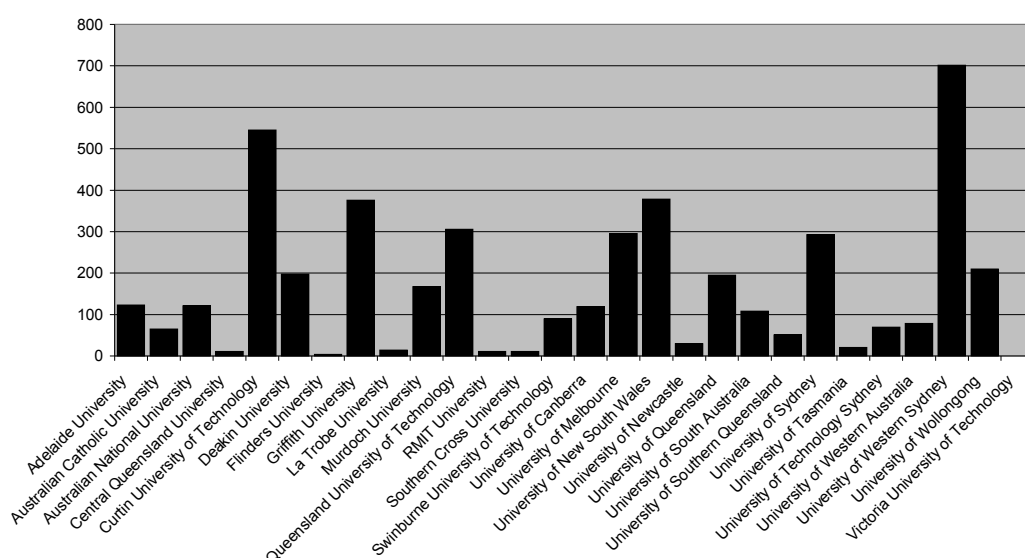
Currently, 26 of the 29 active repositories run an old modified version of Virginia Tech software which is not harvested directly by any robot other than the ADT Program's harvester. Two universities (Western Australia and Curtin) use in-house solutions based around their library catalogue and one other (Melbourne) feeds ETDs from their OAI-PMH-compliant Eprints repository which also holds research articles. The software used by Melbourne was written by Tasmania, which used it as the production repository for six months; however a strange decision by the University Librarian then resulted in a separate installation of the Virginia Tech software. Currently, Tasmania's Eprints mirrors the Virginia Tech repository. Melbourne and Tasmania are the only two universities directly harvested by general and scholarly search engines, though they may be shortly joined by others. This situation is shown in Figure 2.



Content in the Repositories

The theses deposited with the ADT Program can be analyzed by university (Figure 3), and there are a couple of surprising high bars in the chart.

Figure 3 – Total ADT deposits



The five that exceed 300 theses (an arbitrary level) do not all appear to be obvious candidates. The league table of documents and the corresponding number of 2003 research degree completions is shown in Table 1; the rank is based on the 2003 degree completions nationwide.

Table 1 – Ranking of universities with highest ADT deposits

University	Theses in ADT	Completions per year	Completions rank
University of Western Sydney	702	128	18
Curtin University of Technology	545	217	9
University of New South Wales	379	332	4
Griffith University	376	167	13
Queensland University of Technology	306	175	11
<i>Total universities data</i>	<i>4596</i>	<i>6329</i>	<i>39</i>

On analysis, the higher number of theses in these universities appears to derive from a higher rate of investment in digitizing theses which were first accepted and stored in paper form. A later section documents this behavior. This is to be welcomed, but it does not reveal how the universities are adapting to the 21st century and the Internet, or dealing with the imperatives that the open access movement imposes. Accordingly, a more detailed analysis of current deposition rates was undertaken.

The Australian Government collects data on the completions of graduate research degrees (PhD and Master by research), which is publicly available (AVCC, 2005). It is therefore an easy matter to compare this with what is actually happening in the ADT gateway, except that the data is currently available only up to 2003. Accordingly 2003 was chosen as the benchmark year against which to compare ETD deposits. Surprisingly, Table 2 shows that so far on average only an estimated 15% of the current Australian thesis production ends up as ETDs in the ADT system. What is the problem?

Table 2 – Overall success rate in attracting depositions

2003 Completions	6329
------------------	------

Policies

It appears that the problem is the same one that is the bane of research article repositories. If graduates are left to voluntarily contribute an ETD copy of their thesis, few do, and apparently usually at a level at or below 15% (Hajjem *et al*, 2005). Though the extra work to provide an electronic copy is trivial and the thesis is almost always produced electronically anyway, this is just avoidable work and consequently *is* avoided (Swan & Brown, 2005). The reasons for trying to capture *all* theses at the point of submission are worth restating:

- Once the graduate leaves the university, the chance of capturing an electronic copy thereafter is small, due both to human nature (loss of interest) and loss of the CDROM during filing and/or moves (if it is even saved).
- Failure to capture a digital copy of a thesis is forever. A subsequently scanned copy (consisting of page images) is not searchable by search engines or other digital software, and the indexability is lost.
- A key reason is to achieve research impact. A thesis whose full text is not on-line and searchable has a very small chance of being cited or even read. The research may as well not have even been done, except for the graduation.

Accordingly it was decided to analyze the thesis accumulations comparing them with the known completions, and record whether deposition of an ETD was voluntary or mandatory (required in a Rule or Regulation). Twelve Australian universities had mandatory policies at the time of writing as shown in Table 3.

Table 3 – State of mandatory policies in Australia

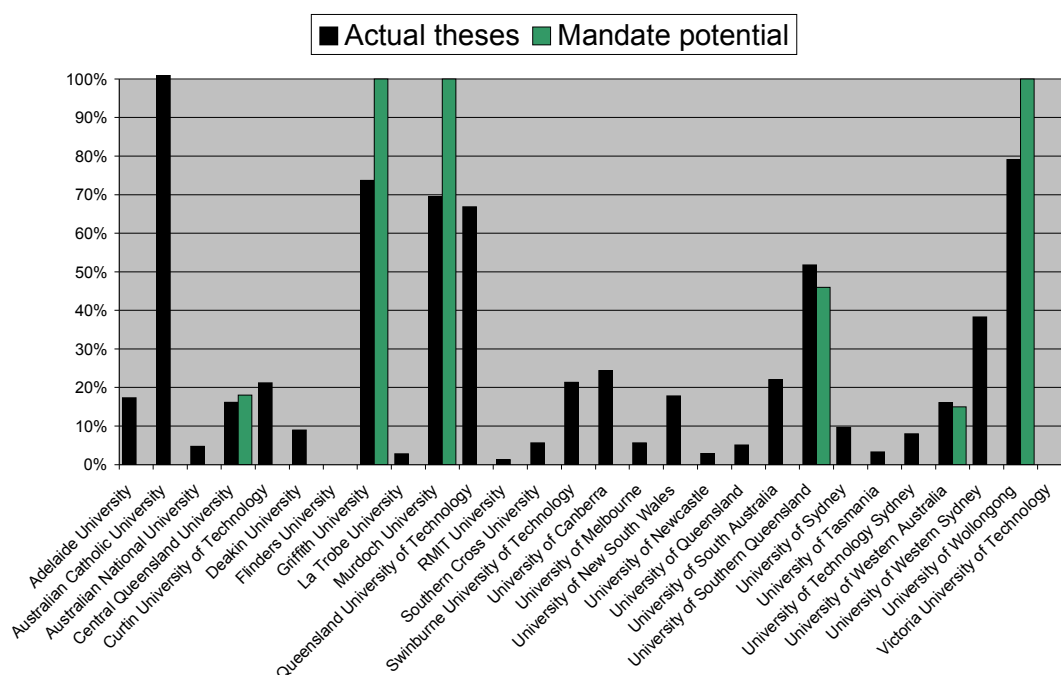
University	Mandatory policy	
Griffith University	Submissions after	Jun 2002
Murdoch University	Submissions after	2003
University of Wollongong	Submissions after	Jun 2003
University of Queensland	Submissions after	2005
University of New South Wales	Submissions after	2005
Queensland University of Technology	Submissions after	Jun 2005
University of Southern Queensland	Enrolments after	2001
Central Queensland University	Enrolments after	2002
University of Western Australia	Enrolments after	2003
Curtin University of Technology	Enrolments after	2005
Swinburne University of Technology	Enrolments after	2005
Monash University	Enrolments after	Jul 2005

It will be noted that in six cases this policy applies to all submissions after a given date and thus have immediate effect. Such policies are rated as having a potential effectiveness of 1.0 (or 100%) after their cut-in date. Another six universities have chosen to have the policy apply to all candidates who enrolled after a given date. This causes the effectiveness to increase with time. Few PhD candidates complete in three years; by four years some but not all of the full-time candidates have completed, and eight years is the normal maximum for part-time candidature but even this may be extended by suspensions. Master results appear more quickly. Accordingly a model-

based effectiveness is assigned to all enrolment policies in force. The modeling is described in a later section.

Figure 4 shows the actual deposits into ADT in 2004, derived from searching the database by university and graduation date = '2004', expressed as a percentage of 2003 completions (graduations). These are derived from the Australian Government's data on PhD and research Master completions for 2003. An average over 1999-2003 was also considered, but the Australian university system is not in steady state. Research degree enrolments are increasing, and hence the 2003 data was considered the best available predictor of the target number of theses that should have been deposited in 2004. [See note at end]

Figure 4 – Deposit success



The results are the same as they are for research articles: a mandatory deposit policy pays off handsomely. Or to put it into a 2 line octosyllabic aphorism:

*'Deposit voluntarily,
See empty repository.'*

All four universities with established policies in this chart (Griffith, Murdoch, Southern Queensland and Wollongong) show that a mandatory policy creates the climate for compliance, and is not overly resisted (Swan & Brown, 2005). Achievement of over 50% compliance in 2004 is good evidence that good policy leads to good outcomes.

There are two apparently anomalous results. It is easiest to discuss the Australian Catholic University first, which appears to be a consequence of very small numbers. This university had only 5 Master and 10 PhD completions in 2003 (a small proportion of the Australian total), and with such small numbers either 2004 completions were higher than in 2003, or persuasion was very effective. It is harder to explain the high deposit levels of the Queensland University of Technology, except that during 2004 they operated a mandatory research article deposit policy, alone of

the Australian universities, and this policy may have influenced graduates to deposit their theses in electronic form in advance of the mandatory thesis deposit policy that they put in place in 2005, or may have caused supervisors to encourage them to do so.

The results are quite clear: a mandatory policy causes deposits to rise to at least 50-80%, compared to the general voluntary policy rate of 5-15%. It is expected that over time a mandatory policy would rise towards 100% deposit, but attitudinal change takes some time.

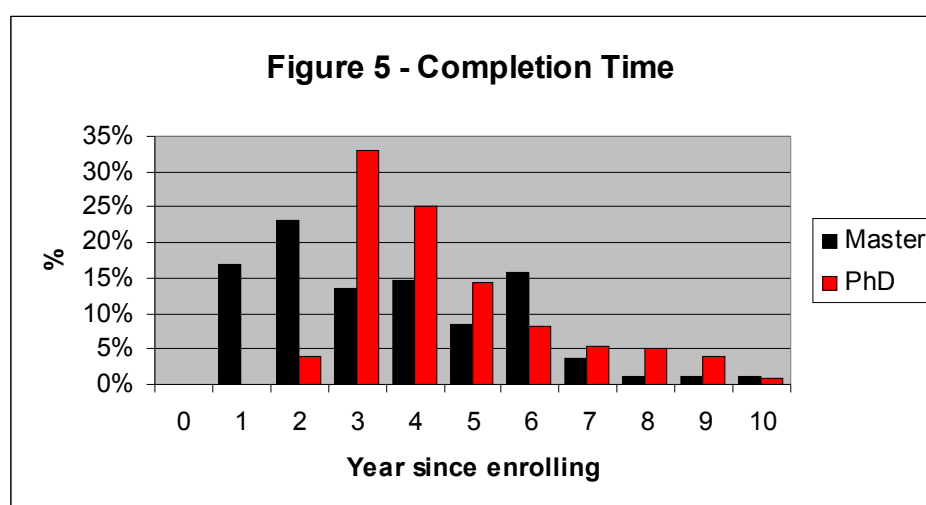
Given this data it is surprising that the Australian Government has not specified that candidates who received a government-funded scholarship (APA) must deposit an electronic copy of their thesis with the degree-granting university. They are aware of the issue, but choose to not become involved in the outcomes of taxpayer funds. Of course, *deposit* does not mean ‘open access’ – it simply means that the ETD describing the research is captured in electronic form, with its metadata. Issues relating to subsequent exploitation by publication or commercial use are not affected.

Modeling of Enrolment-date Policies

In a previous section, the effective implementation rates of the two mandatory policies were mentioned. Here, the figures chosen as the probable target of a mandatory policy are justified.

If a mandatory policy applied to all *thesis submissions* after a given date, this is assumed to bring a 100% target into immediate effect. Such policies are rated as having a potential effectiveness of 1.0 (or 100%) relative to completions, if the decision year is the same as or predates the year under examination (in this paper 2004).

Other universities have chosen to have the policy apply to all candidates who *enrolled* after a given date, usually the date of the decision. This causes the effectiveness to increase with time. Data on the time distribution of completions was not readily available across Australia. Three years of completions (2003-2005 comprising 314 PhDs and 82 Masters) from the University of Tasmania were therefore analyzed and tracked back to their enrolment date (Folvig, 2005). Figure 5 shows the distribution of the elapsed year of completion for these graduates.



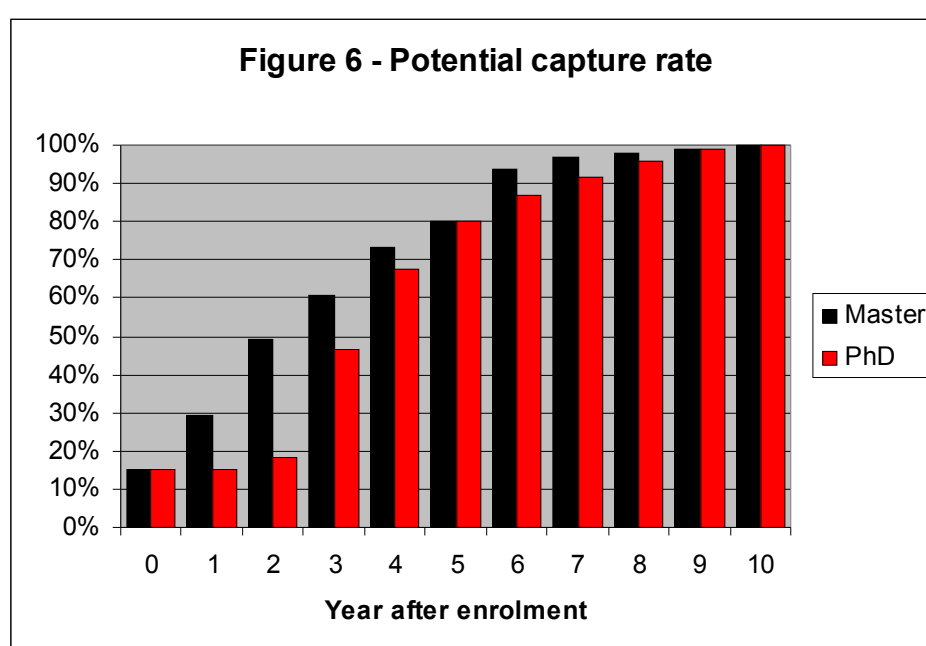
There is no reason to suspect that the University of Tasmania is substantially different from other Australian universities, and accordingly, these distributions are used to model the effect of mandatory policies dated from enrolment.

The available evidence suggests that voluntary policies lead to 15% compliance (the ADT average over Australia, also close to the average for research article repositories, see Hajjem *et al*, 2005). Accordingly a realistic target T for an enrolment-date policy was computed from

$$T = (f_i \times 1.0) + ((1 - f_i) \times 0.15) = 0.15 + (f_i \times 0.85)$$

where f_i is the fraction of candidates who will have completed in the year i

This assumes that all candidates for whom the policy has cut in are recorded at 100%, while the other candidates will voluntarily deposit at 15%. This function is graphed for Master and PhD degrees in Figure 6. It will be seen that an enrolment-date policy does not reach 80% effectiveness (for either kind of degree) until five years after the decision is made!



The value of T for 2004 for all universities with an *enrolment-date policy* is shown in Table 4. If the mandatory enrolment date was after 2004, the university was treated as not having a mandatory policy in 2004, and hence $T = 0$.

Table 4 – Target effect of mandatory policies

University	Policy date	T
Central Queensland University	2002	0.18
Curtin University of Technology	2005	0
Monash University	July 2005	0
University of Southern Queensland	2001	0.46
Swinburne University of Technology	June 2005	0
University of Western Australia	2003	0.15

Costs of delays in instituting fully effective mandatory policies can also be estimated, although it might be better to describe them as *liabilities*, which will be called to account if 100% deposition rate is considered desirable (this should be indisputable) from now or before.

To meet such a target, theses which had not been deposited as electronic copies would need to be scanned and uploaded. It is assumed that a retrospective attempt to capture electronic copies of old theses would have only small success, since academic and graduate record keeping and response rates are likely to be poor. A recent ProQuest study estimates the average time required for scanning at 2hr (Talmacs, 2005), and the average cost at \$55/thesis using a casual rate of \$27.35/hr (ignoring on-costs and profit as would be applied commercially).

- Universities with already effective mandatory policies have no predictable liabilities in meeting such a target.
- Universities with voluntary policies will incur an estimated liability of $(0.85 \times n) \times \$55$

for every year that they do not bring in a mandatory policy, where n is the number of annual completions. For example, for the University of Sydney ($n = 550$ in 2003), this amounts to \$A25 700/year. The choice of a university to use as a hypothetical example is simply that of a large university with a current voluntary policy.

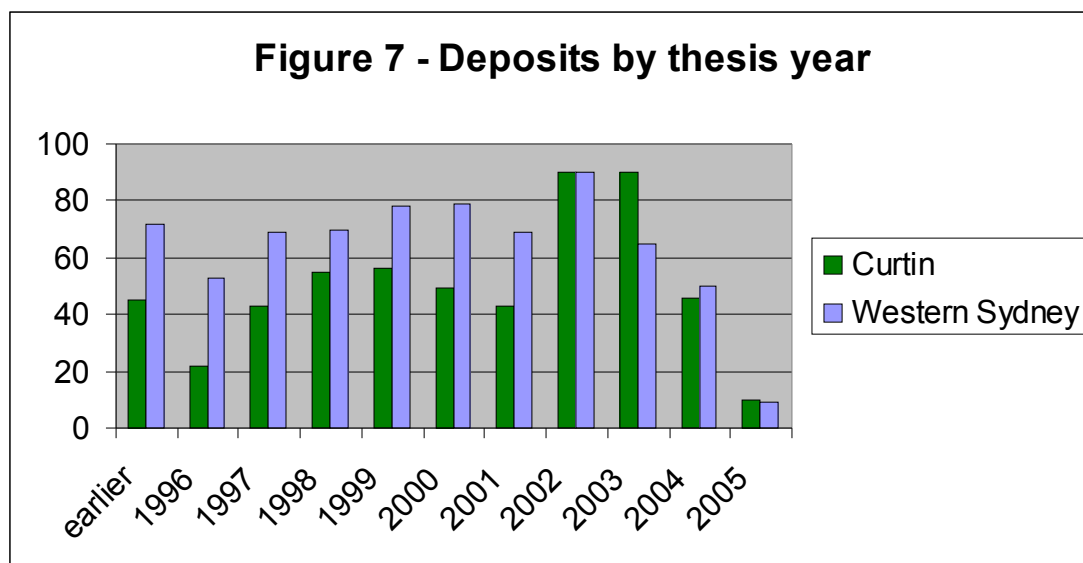
- A university that adopts an enrolment-date policy instead of a submission-date policy incurs a liability through the ramping-up delay which can also be computed using the model in this paper. Taking n_P and n_M as the number of annual completions for PhD and Master degrees respectively, this liability is $((3.06 \times n_M) + (3.84 \times n_P)) \times \55

equivalent to about 3.6 years of a voluntary policy. To again use the University of Sydney as a hypothetical example, this amounts to an unnecessary liability of \$A110 600. These liabilities will almost certainly devolve to the University Library of the institution.

Retrospectivity

It was asserted earlier in the paper that the current thesis counts in ADT were skewed by retrospective deposits. Accordingly, the data for the University of Western Sydney and Curtin University of Technology were analyzed in detail. These are the two largest numerical depositors in the ADT program but 18th and 9th in the PhD-producing universities. The results of the analysis are shown in Figure 7.

Since the ADT Program was only opened up to general deposition in July 2000, all theses from 2000 and before (and some thereafter) are derived from retrospective scanning or serendipitous capture at an earlier time. The fraction is significant. Deposit rates across other universities also suggest that there is a relatively high degree of retrospectivity and a correspondingly lower level of *current* EDT capture.



Consequences

For universities

- 1 Universities which establish an ETD repository seem to be wasting their money if they maintain a voluntary deposit policy. Deposits are poor, running at most at 15% to 20%.
- 2 Mandatory policies pay off handsomely in capturing all theses.
- 3 Mandatory policies established from date of submission are 5-6 years faster in achieving 80% compliance than policies dated from enrolment. All new policies should specify that the policy applies from the date of submission of the final thesis. If there are any quasi-legal quibbles, the Library should specify that it requires *both* paper and electronic copies of the thesis to be deposited under the current rules, which are generally so old that they do not specify format explicitly.

For the Australasian Digital Theses Program

- 4 As the ADT Program is aware, it must now attract the missing universities in Australia, accounting for 12% of the nation's graduate research completions, and the eight New Zealand universities.
- 5 However, a much more important challenge for ADT is the same one that faces operators of institutional research repositories: to increase the deposit rate from an average of 15% of possible thesis production to 100%. This suggests that the ADT Program should adopt a strong advocacy role regarding the adoption of mandatory thesis submission policies.

For the Australian Government

- 6 The Australian Government could accelerate the development of mandatory submission policies by a simple administrative action, costing nothing. If the guidelines for Australian Postgraduate Awards (APAs) were slightly amended to require a completing graduate to deposit both a paper copy and an electronic copy of his or her thesis with the university from which they graduate, the deposit rate would jump immediately, and universities would rapidly insist on

the same action by their graduates not in receipt of these awards. This is a simple consequence of the 'Backing Australia's Ability' policy.

Acknowledgments

The author would like to thank:

- Ian Mitchell in Research Services and Vanessa Folvig in the Graduate Research Unit of the University of Tasmania, for quickly providing requested data;
- Tony Carneglutti and Kerrie Talmacs at the University of New South Wales for ADT Program data, and for the Program gateway itself – the subject of this research.

This work was carried out in the School of Computing, University of Tasmania.

References

- AuseAccess (2005). Australian Open Access Wiki available at <http://leven.comp.utas.edu.au/AuseAccess/>.
- Australian Digital Theses Program (2005). Gateway available at <http://adt.caul.edu.au/>.
- Australian Vice-Chancellors Committee (2005). *HERDC Time Series Data 1992-2003*. Available at <http://www.avcc.edu.au/documents/publications/stats/HERDCTimeSeriesData1992-2003.xls>.
- Folvig V, University of Tasmania (2005). Unpublished communication regarding completion data at the University of Tasmania.
- Hajjem C, Harnad S, Gingras Y (2005). Ten-Year Cross-Disciplinary Analysis of the Growth of Open Access and Its Effect on Research Citation Impact. Paper in preparation, accessed at <http://www.ecs.soton.ac.uk/~harnad/.Tmp/ieee5.doc>.
- Swan A and Brown S. (2005) Open access self-archiving: An author study. [http://www.keyperspectives.co.uk/openaccessarchive/reports/Open%20Access%20II%20\(author%20survey%20on%20self%20archiving\)%202005.pdf](http://www.keyperspectives.co.uk/openaccessarchive/reports/Open%20Access%20II%20(author%20survey%20on%20self%20archiving)%202005.pdf)
<http://eprints.ecs.soton.ac.uk/10999/>
<http://cogprints.org/4385/>
http://www.jisc.ac.uk/uploaded_documents/Open%20Access%20Self%20Archiving-an%20author%20study.pdf
- Swan A and Brown S (2004). ISC/OSI Journal Authors Survey Report. <http://eprints.ecs.soton.ac.uk/11002/>
<http://cogprints.org/4125/>
http://www.jisc.ac.uk/uploaded_documents/JISCOARReport1.pdf
- Talmacs K, University of New South Wales (2005). Unpublished communications regarding costs of thesis scanning.

[3041 words excluding references]

Note to referees: Australian Government data for 2004 completions is expected to become available in December 2005 or January 2006. If this data becomes available before finalization of the paper (desirable), the tables and the paper will be updated to compare 2004 completions (instead of 2003) with 2004 ADT data.

Biographical Sketch

Arthur Sale PhD FACS FIEAust is Professor of Computing Research at the University of Tasmania and Research Coordinator of Computing. Prior to that he was Pro Vice-Chancellor of Information Services and a University Senior Executive. Dr Sale has been described as the 'archevangelist of Open Access in Australia' (Stevan Harnad, Canada) and a 'Living National Treasure' (John Hughes, UTS). He received the ANCAAC Award for best paper in Australia in ICT in 2001 and the ICT Industry Award for contribution to the ICT Industry in 2004. He researches in bioinformatics, mobile computing and open access.

